DOCUMENT RESUME

ED 276 770 TM 860 722

AUTHOR Webster, William J.
TITLE The Practice of Evaluation in the Public Schools.
PUB DATE 16 May 86
NOTE 38p.; Paper presented at the Minnesota Evaluation Conference (May 16, 1986).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Accountability; Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Flow Charts; Formative Evaluation; *National Surveys; *Program Evaluation; *Public Schools; *Resource Allocation; School Districts; School Organization; School Surveys; *Summative Evaluation
IDENTIFIERS *Dallas Independent School District TX

ABSTRACT

This paper examines trends in the use of research and evaluation resources in the Dallas (Texas) Independent School District (DISD) and in other school districts throughout the country. First, the organization and functions of the Department of Research, Evaluation, and Audit of the DISD are presented and compared to those of 7 years ago. Second, the results of a 1983-84 survey of the 90 largest school districts to recent evaluation funding and activities versus funding and activities reported in a previous survey. The results of these surveys, as well as the Dallas experience, are used to illustrate the changing emphases of public school evaluation over the past several decades. (JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED276770

THE PRACTICE OF EVALUATION IN THE PUBLIC SCHOOLS[1]

William J. Webster
Dallas Independent School District

This paper examines trends in the use of research and evaluation resources in the Dallas Independent School District (DISD) and in other school districts throughout the country. First, the organization and functions of the Department of Research, Evaluation, and Audit of the DISD are presented and compared to those of seven years ago. Second, the results of a 1983-84 survey of the ninety largest school districts in North America are discussed relative to recent evaluation funding and activities versus funding and activities reported in a previous survey (Webster and Stufflebeam, 1978). The results of these two surveys, as well as the Dallas experience, are used to illustrate the changing emphases of public school evaluation over the past several decades.

Organization - DISD

The Dallas Independent School District's Department of Research, Evaluation, and Audit is responsible for all data collection, analysis, and reporting activities of the DISD. Its major purposes are to provide information for use in decision-making and to serve as an accountability agent to the District's General and Deputy Superintendents, the Board of Education, and the general public. The Department reports directly to the General Superintendent and, as a result, is independent of all program development and operations departments. This independence is essential if the accuracy and credibility of reports and data are to be assured.[2]

---

[1]Paper presented at the Minnesota Evaluation Conference, May 16, 1986.

[2]A recent survey of twenty-two large-city evaluation departments showed that seven reported directly to the Superintendent, three were housed within an administrative division, five were housed within an instructional division, six were housed within a support division, and one was housed within a planning division (Taylor and Casserly, 1985).

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
W. Webster
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM 860 722

A somewhat unique feature of the DISD's reporting arrangement is that all evaluation reports are presented to and discussed by a permanent sub-committee of the Board of Education. This committee, the Program Evaluation Committee, meets monthly and often spends many hours discussing the implications of various evaluation reports. The DISD's Department of Research, Evaluation, and Audit also reports to the Federal Court three times per year relative to the District's efforts and outcomes in implementing a Federal Desegregation Court Order.

Figure 1 presents an organization chart for the Department for 1986-87. The specific functions of the Department are discussed in the next section of this paper.

See Figure 1

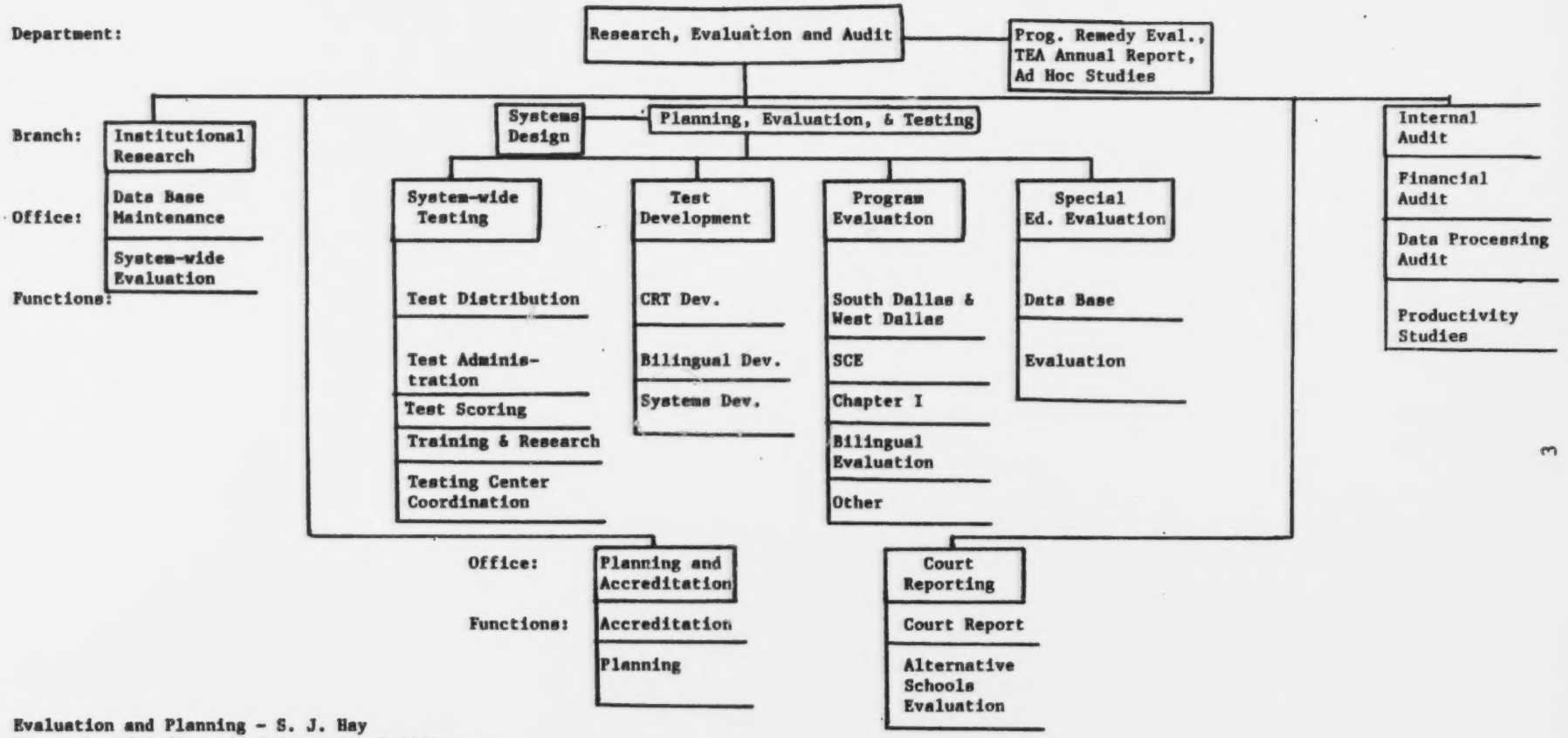
Functions - DISD

Research and evaluation activities in the DISD are more precisely defined and important than at any time in the sixteen-year history of the Department. Every research or evaluation project that is undertaken is done because the District, or some specific agency, acutely needs the data. This is partially the result of reduced resources and partly the result of a managerial philosophy that stresses accountability at all levels of the organization.

The DISD utilizes the CIPP Model in conducting its research and evaluation activities (Stufflebeam, et al, 1971). Figure 2 provides a schematic that depicts how the model operates in the DISD. Discussion of Figure 2 is used in this paper as a vehicle to explain the different emphases of the Department today versus seven years ago. For a detailed discussion of the model see Webster (1975).

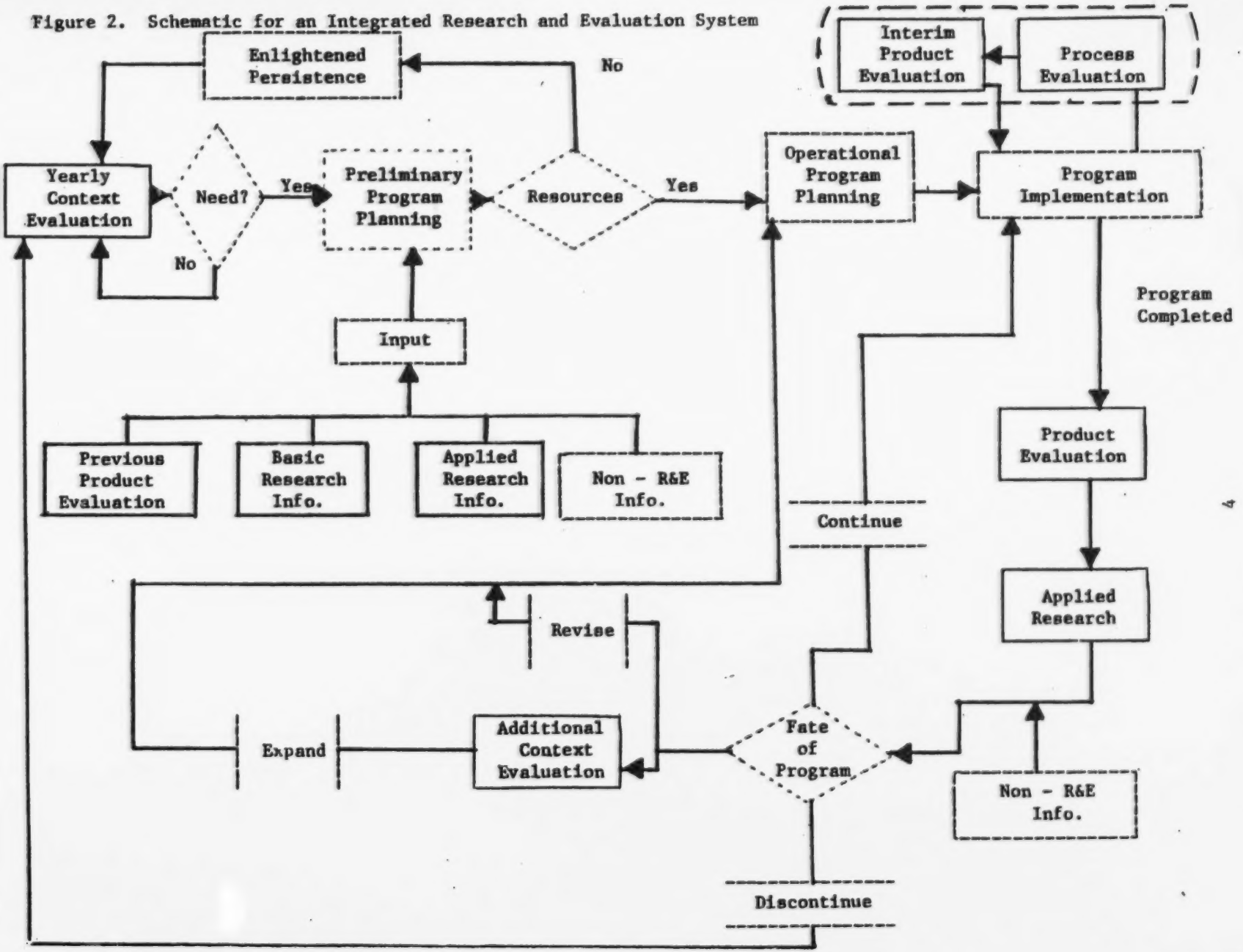
See Figure 2

Figure 1. Organization Chart for the Department of Research, Evaluation, and Audit - 1986-87



Evaluation and Planning - S. J. Hay
Research and Audit - Administration Building
Institutional Research - Administration Building
Internal Audit - Administration Building

Figure 2. Schematic for an Integrated Research and Evaluation System



A prerequisite to improvement or, for that matter, accountability, must be a knowledge of existing performance levels. Thus, the backbone of any renewal system must be a comprehensive context evaluation system. Stufflebeam, et al (1971), defines context evaluation as the provision of baseline information that includes the environment of interest, describes desired and actual conditions pertaining to the environment, identifies unmet needs and unused opportunities, and diagnoses the problems that prevent needs from being met and opportunities from being used. Context evaluation is often referred to by Stufflebeam and others as needs assessment.

The heart of the Department of Research, Evaluation, and Audit's evaluation activities is its context evaluation system. This system is based upon a ten-year longitudinal data base that enables important demographic and achievement data on individual students to be followed over time. The system is designed in such a manner that, on a systemwide basis, this year's context evaluation becomes next year's product evaluation. For example, the context evaluation for the 1985-86 school year is the product evaluation for the 1984-85 school year. The existence of the aforementioned data base permits ad hoc data inquiries that, without the data base, would be impossible or, at best, extremely expensive (Webster and Schuhmacher, 1973).

As of this date, the District's data base provides for context evaluation at the student, classroom, school, subdistrict, and district levels. All evaluation activities initiated in the District are far more school-oriented than they were during the decade of the seventies. The Dallas evaluation system was originally designed to serve primarily the information needs of top management and program management. Through the years, it has expanded to serve the information needs of all levels in the institution. This change in emphasis has tracked the gradual disappearance of specific program management and the accompanying pressure for accountability to be centered on the principal at the local school level.

Once the context evaluation system has identified needs, decision-makers must prioritize those needs and focus on reducing the discrepancy between desired and existing conditions by establishing goals

for those needs that receive highest priority. It is at this point in time that input evaluation information is used. Stufflebeam, et al (1971), defines input evaluation as the provision of information for determining methods of resource utilization for accomplishing program goals. In a functioning evaluation system, there are four major sources of input information:

1. previous product evaluation information,
2. basic research information,
3. applied research information, and
4. non-research and evaluation information.

Product evaluation information concerns the extent to which specific project or program goals are achieved. When product evaluation information is available relative to a given program with goals similar to those identified in response to context evaluation information, that information provides useful input to decision-makers in determining the probability that the program would reduce the identified discrepancy between desired and existing conditions.

Basic research information pertains to information about fundamental relationships that affect student learning. Before making a decision to implement a given program, decision-makers should be apprised of the extent to which that program is or is not consistent with the principles established by basic research in learning.

Applied research information concerns the interaction between student characteristics, teacher characteristics, and instructional systems. Applied research differs from basic research in that the information provided is more closely related to specific decisions in an applied educational setting. Decision-makers need information relative to the types of students that function best in given instructional systems implemented by specific types of teachers. Applied research must be implemented in the public schools by public school research departments. Basic research should probably be implemented in the universities in conjunction with the public schools since many local boards of education are probably too impatient for results and are unlikely to wait for a long-term, systematic, longitudinal, basic research program to pay off.

Finally, non-research and evaluation information must enter into most educational decisions. Such information as capabilities of staff members,

costs, political feasibility of program implementation, and existing facilities must be taken into account.

After the collection of relevant input information feeding a preliminary program planning stage, decision-makers determine whether or not sufficient resources are available to make the desired changes. Quite often, sufficient resources are not available and some compromise is necessary. In many cases, the lack of resources is not limited to the realm of cost and political feasibility, but rather stems from an insufficient base of knowledge. Thus, educators are often in the position of having sufficient material resources but insufficient information resources.

If sufficient material resources are not available, the system may have to exist for some period of time in a state of enlightened persistence. Periodic context evaluation will continue to highlight the extent of discrepancy between that which is desired and that which exists. If the problem results from insufficient information resources, programs are often implemented without sufficient support data, and an information base is built through a series of systematic evaluation and applied research studies.

If sufficient material and information resources are available, or if sufficient material and minimal information resources are available, the program planning phase is entered. The evaluator's role in program planning involves making all relevant input information available to program planners and ensuring that stated program objectives are measurable. Out of the program planning sessions, the evaluator develops a detailed program evaluation design specifying the criteria by which the program will be judged. The development of this evaluation design must necessarily involve continuous interaction between program personnel and the evaluator in order to produce maximally effective information. Obviously, the evaluator must be independent of program management to ensure the optimum objectivity of evaluation results.

Once the program implementation phase is entered, the evaluator's role becomes one of providing continuous formative evaluation reports relative to program implementation. These reports fall primarily into two categories -- process evaluation and interim product evaluation. Process

evaluation has three major objectives: (1) the detection or prediction of defects in procedural design or its implementation during program implementation stages, (2) the provision of information for programmed decisions, and (3) the maintenance of a record of the implementation procedure as it occurs (Stufflebeam, et al, 1971). Thus, process evaluation information keeps program management informed of the extent to which program implementation conforms to specifications and, from an evaluation standpoint, guards against the evaluation of a fictitious event.

Much of the process evaluation which was at one time implemented by DISD evaluation personnel is now implemented by subdistrict and curriculum/instruction staff. This is consistent with the accountability emphasis that is currently the philosophy of District management. Since process evaluation is extremely expensive, many of the cutbacks in research and evaluation activities over the past few years have been in the area of process evaluation.

Interim product evaluation provides periodic feedback to program management relative to the attainment of specific sub-objectives during the implementation phase. Thus, process and interim product evaluation reports inform program management as to implementation and goal attainment levels while program adjustments are still feasible. Program evaluation, relative to specific developmental projects, has also been reduced in the DISD over the past five years. This is largely because the number of developmental projects has been reduced.

Upon completion of a given cycle of program implementation, a summative product evaluation report is prepared. This report generally addresses three areas of concern:

1. the extent to which program objectives were achieved relative to some specific set of criteria,

2. the extent to which system objectives were achieved relative to alternative instructional strategies, and

3. the cost-effectiveness of the program relative to alternative instructional strategies.

It should be obvious that if the principal objective of an evaluation is program improvement, information relative to these three areas of concern must be interpreted in light of process and interim product evaluation

data. Such an interpretation is not nearly so important in an accountability context. A detailed discussion of an operational project evaluation process can be found in Webster (1973).

Most evaluation systems stop at the provision of product evaluation information. These data generally bear upon the performance of different groups of students under varying treatment configurations. Unfortunately, as a result, most product evaluation reports have generally focused on the search for single best treatments for all learners, i.e., main effects. In order to provide needed information for educational decision-makers, applied research studies involving the systematic investigation of aptitude-treatment and trait-trait interactions must be undertaken. Such studies would be expected to provide important information relative to replicable relationships between student, teacher, and program characteristics (Bracht, 1969; Cronbach and Snow, 1969; Salomon, 1972; Webster and Mendro, 1974). The DISD no longer has a formal applied research program, although some applied research is conducted through the various program evaluations.

Once context, input, process, and product evaluation information, as well as applied research data are available, non-research and evaluation information once more is brought to bear upon the decision-making process. It would be naive to expect educational decisions to be made purely on the basis of research and evaluation data. Once again, information such as the absolute program costs, capabilities of program staff members, political feasibility of program implementation, and existing facilities and resources must be considered by decision-makers in rendering a final decision.

In determining the fate of a given program, four primary choices are available to decision-makers. First, they can choose to continue the program in its current setting. If this alternative is chosen, the summative product evaluation report and the applied research data become the context evaluation information for the next implementation phase and program implementation commences. This alternative generally occurs when decisions are to be made on the basis of longitudinal studies, i.e., where it is expected that results will not be in evidence after a relatively short implementation period.

A second alternative is to discontinue the program. This is usually done after product evaluation studies demonstrate the failure of the program to meet its objectives or in those cases where the program is simply not cost-effective. (Failure to meet objectives is often a necessary but not sufficient condition for program discontinuation.) Once a program is discontinued, the system returns to context evaluation and once again applies the needs assessment and orientation phases.

If the product evaluation and applied research information are favorable, and it is practically and politically feasible, the program may be expanded to serve additional students. Prior to making the decision to expand the program, additional context evaluation information must be examined to determine if similar needs exist in other settings. If such needs are demonstrated, then the program may be expanded to other settings, and the program planning stage entered to extend the program implementation. Other settings eligible for program expansion may include entire schools or specific sub-populations (e.g., highly motivated students) as indicated by applied research data. If such needs do not exist, the program is continued with the original target population or a reduced target population based on the results of the applied research studies. The extent of continued evaluation to be implemented under either the expansion or continuation alternative is determined by decision-makers under advisement from evaluation personnel.

A fourth alternative involves program revision. Much program revision should be accomplished on the basis of process and interim product evaluation reports. Often, however, summative product evaluation and applied research reports highlight weaknesses in portions of programs that would otherwise appear to be functional. In this instance, the summative product evaluation and applied research reports become the context evaluation information for the next program planning cycle.

Figure 3 outlines the proportion of funds that were spent by the DISD for various research and evaluation activities in 1978, 1983, and those proposed for 1986-87. Study of Figure 3 suggests a trend between 1978 and 1983 toward more testing and test development and less input evaluation, process evaluation, and applied research. Ad hoc data requests were also up significantly. Process evaluation was carried out primarily by curriculum and instruction personnel with aid and training from evaluation

personnel. The ad hoc requests for information were up five hundred percent over 1978, but the District's longitudinal data base enabled this increased demand to be handled very efficiently. The increase in planning services was largely the result of greater use of evaluation data by the District, which causes evaluation personnel to be more involved in providing input to program designers and decision-makers. This increased involvement was particularly true at upper decision-making levels. The development of instructional use of technology was discontinued in 1980.[3]

The 1986-87 proposed budget for research and evaluation activities suggests some interesting trends. Testing continues to claim a larger percentage of the budget. This trend will continue as the District implements its criterion-referenced testing system, grades K-12.[4] Test development costs will continue to decline as item pools become available. The context evaluation system produces more information than ever, but is becoming increasingly efficient to the point where much of the resources spent on this function are spent on developing new systems. Input evaluation is increasing as the research data have more influence on decisions. Process evaluation is increasing as the Federal Court and the Superintendent require additional reliable information on what is going on in the schools. Finally, financial audit has been added as an additional responsibility of the Department.[5]

See Figure 3

---

[3] In 1978, the District as a whole spent more than thirteen million dollars on curriculum development. For 1986-87, the comparable figure will be less than one million dollars.

[4] The criterion-referenced testing program is projected to have operating costs in the neighborhood of $750,000 per year. However, this system will eventually replace teacher-made final examinations, will be used in assigning final student grades, and will be used in teacher evaluation as well as school-wide performance evaluation.

[5] Montgomery County, Maryland, also houses the financial audit function within the evaluation department.

Figure 3. Proportion of Research and Evaluation Resources Spent on Various Evaluation Activities, DISD, 1978 and 1983

Funds:	1978	1983	Proposed 1986-87
Local/State Fund	$ 1,879,491.	$ 1,761,396.	$ 3,319,884.
Federal Funds	1,483,257.	515,000.	490,500.
	$3,362,748.	$ 2,276,396.	$ 3,810,384.
Personnel:[1]			
Professional	111	41	36
Support	43	17	16
Activities[2]			
Management	5	4	5
Testing	15	23	27
Instrument Development	4	18	10
Context Evaluation	8	12	8
Input Evaluation	5	1	4
Process Evaluation	16	5	7
Product Evaluation	17	21	14
Applied Research	10	1	2
Basic Research	1	0	0
Data Processing	5	5	5
Planning Services	1	3	3
Ad Hoc Information	1	5	5
Research Consultation	1	1	1
Proposal Development	1	1	1
Financial Audit	0	0	8
Technology Usage (Development)	10	0	0

[1] For 1986-87 there are an additional five professional and one support personnel in Internal Audit. In addition, the Court has ordered a comprehensive evaluation of Six Learning Centers that will require a staff of five professional and one support personnel.

[2] Figure 5 defines the various functions except for financial audit and technology usage (development). Financial audit involves the systematic audit of the District's financial records. Technology usage/development involves the development of learning systems utilizing technology. This is not really an evaluation function but often is associated with evaluation departments because of their expertise in technology. This endeavor is now correctly placed in the Instructional Support Division.

From a funding standpoint, local and state funds decreased by $118,095 between 1978 and 1983. This represents about a six percent decrease. Over this same period of time, the DISD's total budget increased by some 55%. Thus, evaluation resources did not keep pace with the increase in District resources. Federal evaluation resources fell by some 65%. This reflected the decline in Federal monies for all activities, particularly research and development activities.

For 1986-87, the proposed budget has surpassed the 1978 level. Most of the increases from 1983 are in testing and the financial audit function. Only 57% of the Department's budget is personnel. Thus, although Department personnel have been cut 66.2% since 1978, increased computerization has helped maintain high levels of production. As mentioned earlier, the amount of process and product evaluation has decreased dramatically since 1978. All development activities within the department have ceased. (Development activities have also been severely curtailed throughout the District.) Federal funds continue to decline. Figure 4 presents the proposed program budget for the Department for 1986-87. Study of this figure should provide the reader with an accurate picture of the range of evaluation activities conducted.

See Figure 4

The Practice of Evaluation - The National Perspective

The field of educational evaluation developed dramatically during the decade of the 1960's, influenced in part by the seminal theories of Cronbach (1963), Stufflebeam (1966), Scriven (1967), and Stake (1967), and stimulated by the evaluation requirements of the "great society" programs launched in 1965 as well as the national accountability movement of the 1970's. The practice of educational evaluation in the public schools expanded equally as dramatically during the decades of the 60's and 70's. By 1976, approximately 350 of the nation's 15,000 school districts maintained formal research and development units (NIE Databook, 1976).

Two surveys (Lyon, et al., 1978; Webster and Stufflebeam, 1978) suggested the nature of activities conducted by evaluation units during the seventies. The Lyon survey, including responses from 227 of 319 districts in the United States with student populations of 10,000 or more, concluded

Figure 4. Program Statement: Department of Research, Evaluation, and Audit, 1986-87

		G	O
1.0	Manage the research, evaluation, and audit-functions of the District through prudential planning, organization, staffing, and operation.		
1.1	Establish departmental objectives and secure necessary material (budget) and personnel resources to accomplish these objectives.		
1.1.1	Develop departmental objectives and associated budget	$ 6,000	
1.1.2	Develop and review detailed evaluation and audit plans	35,500	$ 7,000
1.1.3	Quality control designs and reports	15,000	
1.1.4	Recruit and hire qualified personnel	5,000	
1.1.5	Implement an improved employee evaluation system	19,500	3,600
1.2	Provide for the daily operation of the various departmental offices.		
1.2.1	Manage the Department	29,664	
1.2.2	Control and amend the budget	5,000	
1.2.3	Review proposals for outside research	2,500	
1.2.4	Edit major departmental publications	4,000	
1.2.5	Conduct necessary daily interaction with internal and external agencies	12,500	3,000
1.2.6	Operate the S.J. Hay building	30,000	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

		G	O
	1.2.7 Print department final reports	45,000	
2.0	Provide technical support to District desegregation activities.		
2.1	Produce the annual planning guide	6,000	
2.2	Provide demographic data related to desegregation alternatives (input evaluation)	7,500	
2.3	Provide liaison between the District and the trial lawyer, as well as the Court and Court Auditor, on issues that are being litigated	4,000	
2.4	Provide ad hoc services on desegregation	3,000	
2.5	Contract with the Court Auditor to provide auditing services to the Court	155,000	
3.0	Provide valid data for decision-making and accountability purposes.		
3.1	Provide ad hoc research services at the request of the General and/or Deputy Superintendent(s)	3,000	
3.2	Oversee the contract for the study of the Hispanic dropout	2,500	
	3.2.1 Dropout contract for 1986-87	100,000	
3.3	Coordinate data collection and generate the Annual Performance Report required by H.B. 72 and subsequent TEA interpretations	30,000	
3.4	Provide the district's input evaluation function. Recent activities include student grading, facilities, and early childhood education	6,500	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

		G	O
3.5	Keep abreast of current trends and developments in the social science literature	20,000	
3.6	Interpret results of evaluation and research studies to ensure that they are optimally utilized	34,000	4,200
4.0	Coordinate the monthly meeting of the Program Evaluation Committee of the Board of Education.	12,050	
5.0	Provide objective, valid context evaluation data for accounting to District decision-makers, the Federal Court, and the public relative to the District's primary product: the achievement and accomplishments of its students.		
5.1	Maintain the computerized longitudinal data files for all system-wide tests and student demographic data		
5.1.1	Continue development of data base software	10,500	
5.1.2	Produce Individual School Achievement Reports	19,500	
5.1.3	Produce Achievement Indices and Objective Attainment Analyses	15,000	
5.1.4	Produce Performance Profiles (comprehensive reports on school achievement to area administrators)	8,000	
5.1.5	Produce Quasi and Cohort Longitudinal Reports	12,550	
5.1.6	Produce Reports on the Survey Tests of Essential Elements/ Learner Standards Reports (Criterion-Referenced Tests)	30,250	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

			G	O
	5.1.7	Produce Enrollment Trends and Projections	4,500	
	5.1.8	Provide Accreditation Support (Objective Attainment Analyses, Skills Analyses, School Improvement Objectives Worksheets, School Improvement Planning Worksheets)	18,850	
	5.1.9	Provide technical support to Court-ordered reporting requirements	5,250	
	5.1.10	Produce Reconfigured School Reports (Fall)	12,500	
	5.1.11	Produce ad hoc reports from the data base (100 per year)	20,000	
5.2	Provide special reports and/or services to high interest topics:			
	5.2.1	Graduate Follow-Up	4,000	
	5.2.2	SAT/ACT/PSAT	1,500	
	5.2.3	Dropout Studies	12,500	
	5.2.4	Montessori	5,250	
5.3	Continue the development and analysis of the District's Effective Schools Methodology including the modification of the methodology to include criterion-referenced test results		18,250	
5.4	Further develop and pilot the methodology necessary to link student criterion-referenced test scores to teachers		7,500	
5.5	Pilot the system that utilizes criterion-referenced test results in assigning student performance grades. This		14,250	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

		G	O
	system must stress quality control at every point as well as rapid turnaround.		
5.5.1	Study the feasibility of posting criterion-referenced tests scores directly to the data base	3,250	
5.6	Norm the criterion-referenced tests produced during 1985-86	4,000	
5.7	Study the feasibility of replacing the Tests of Achievement and Proficiency, grades 9-11, with the secondary criterion-referenced tests for 1987-88	5,000	
6.0	Provide the District with objective information for planning and decision-making through the implementation of the system-wide testing program.		
6.1	Purchase, distribute, and score the following tests and provide individual student reports:		
6.1.1	Iowa Tests of Basic Skills Grades K-8	175,000	
6.1.2	Tests of Achievement and Proficiency (Grades 9-11)	48,500	
6.1.3	Language Assessment Scale, Grades Pre K-12	37,750	
6.1.4	CTBS-Espanol, Grades 1-11	12,500	
6.1.5	TEAMS, Grades 1, 3, 5, 7, 9, F11, S11, F12, S12	20,500	
6.1.6	Survey Tests of Essential Elements/Learner Standards, Grades 1-12	45,000	
6.1.7	Spanish Version Survey Tests of Essential Elements/ Learner Standards, Grades 1-3	12,250	
6.1.8	Special ITBS, TAP, and CoGat for Talented and Gifted Program	10,000	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

		G	O
6.2	Conduct special testing studies in conjunction with TEA and Testing companies.		
	6.2.1 Organize, study, receive, distribute, and return materials within study timeline	10,250	
	6.2.2 Conduct training, monitor and quality control study	9,500	
6.3	Design and develop computer programs for scoring, reporting and distributing all tests (the volume of tests being scored is increasing at the rate of 100,000 per year).	17,500	
6.4	Design and print machine-scorable booklets (K-2) and machine-scorable answer sheets (3-12) for criterion-referenced survey tests and control documents for (K-12).	43,750	
6.5	Conduct in-service training with building test coordinators (K-12) in proper implementation of Districtwide testing program.	50,000	
6.6	Conduct administrative staff, teacher, and parent in-service programs on the purposes, utilization, and interpretation of tests.	5,250	
6.7	Conduct necessary research to assure the optimum validity of the testing program to include test standardization, levels articulation, effects of off-level testing, and validity, and comparability of norms.	15,000	
6.8	Develop and print administrative manuals, test coordinators handbooks and parent pamphlets.	20,000	
6.9	Insure the validity of 1987 Systemwide achievement data through systematic monitoring of test administration procedures and through valid test equating procedures.	50,000	
6.10	Purchase packaging materials for tests and provide part-time temporary assistance at Service Center.	6,250	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

		G	O
6.11	Design and operate a District Testing Center at S. J. Hay.	60,000	
7.0	Design and validate new assessment criteria which validly measure the effects of instruction while diminishing race and class bias.		
7.1	Revise and validate existing criterion-referenced survey tests and expand number of items available for tests in grades 7-12. This will include 11 tests at 7-8 [Communications (3), Math (3), Social Studies (2), Science (3), and selected courses by semester at the high school level, 9-12 Math (6), Science (5), Social Studies (6), Communications (5)].	145,220	
7.2	Revise and validate existing criterion-referenced survey tests and expand number of items available for tests in grades 1-6 (Language Arts, Mathematics, Social Studies, and Science).	95,500	
7.3	Revise and validate existing Spanish editions of the Criterion-referenced Survey and classroom Tests, 1-3.	24,250	
7.4	Print survey and classroom criterion-referenced tests.	404,600	
7.5	Expand local school micro-computer test scoring systems in West and South Dallas Centers.	8,000	
7.6	Develop and validate criterion-referenced survey tests for additional courses. Courses include Correlated Language Arts (Grade 10), Reading Improvement (Grade 10), Sociology, Psychology.	35,000	
7.7	Continue research on prototypic item banking system including appropriate taxonomic structure.	10,000	
7.8	Research and implement appropriate equating methodologies.	18,800	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

		G	O
7.9	Develop and pilot-test kindergarten survey test.	7,500	
7.10	Develop kindergarten Classroom Test and Classroom Tests for other subject areas or courses that are identified as priorities by teachers who work on them during the Summer 1986.	7,500	
8.0	Redesign the accreditation model to reflect reduced monitors and new requirements in H.B. 72 and its interpretation.		
8.1	Provide data and forms needed for schools to conduct their needs assessment.	25,000	
8.2	Provide technical assistance to schools and administrators in the design and implementation of the planning/accountability model.	37,500	
8.3	Assist in monitoring school plans and in working with Operations and Instructional Support personnel to provide needed data.	125,680	
8.4	Report results to the District and State.	12,500	
9.0	Provide objective and valid process and product evaluation data for decision-making.		
9.1	Provide accurate and timely evaluative information for the Bilingual Education programs.		
9.1.1	Develop and maintain systemwide database for Bilingual Education	8,500	
9.1.2	Evaluate Districtwide BE/ESL general programs	40,000	
9.1.3	Evaluate BE/ESL summer school	5,000	
9.1.4	Evaluate Bilingual Education pre-school summer school	5,000	
9.2	Provide information as required by the Court Order for the Learning Centers.		

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

			G	O
	9.2.1	Develop and maintain database for the Learning Centers	5,250	
	9.2.2	Implement the Court-Ordered evaluation	208,965	
9.3		Provide state mandated evaluations to the Texas Education Association for the following programs:		
	9.3.1	Reading Improvement - Grade 7-8	25,750	
	9.3.2	Reading Improvement - Grade 9	25,750	
	9.3.3	School Based Tutoring - Grades 7-8	22,750	
	9.3.4	Basic Skills Core Program	5,250	
	9.3.5	Correlated Language Arts	5,250	
	9.3.6	Fundamentals of Mathematics	5,250	
	9.3.7	Seventh Six-Weeks Summer School Grades 7-8	12,750	
9.4		Provide administration and Board of Education information as to the effectiveness of the Seventh Six-Weeks Summer School program, Grades 1-6.		
	9.4.1	Design, develop, and administer pre and post-testing on targeted learner standards	4,250	
	9.4.2	Provide final report to the administration	12,750	
9.5		Provide mandated evaluation services and reports for the following extra-district funded programs.		
	9.5.1	Chapter I programs including: Basic Skills, Writing To Read, Community Involvement, Institutions for the Neglected and Delinquent, Catholic Dioceses, Summer School	5,000	256,700

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

			G	O
	9.5.2	Special Education including: Context report and other projects as identified by the administration and/or management	1,000	138,000
	9.5.3	Chapter II programs	1,000	14,000
	9.5.4	Chapter I Migrant	1,000	14,000
	9.5.5	Magnet School Improvement	8,800	
	9.5.6	Saxon Mathematics Program	8,800	
	9.5.7	Miscellaneous Programs as specified in proposals	10,000	40,000
9.6		Provide data and services on request to schools, teachers, administration, Board of Education and outside organizations.		
	9.6.1	Talented and Gifted Program	8,500	
	9.6.2	Unique requests for rosters or analyses from schools	8,000	
	9.6.3	School desegregation monitoring	4,000	
	9.6.4	Drug Usage Studies	2,500	
	9.6.5	Ad hoc studies as requested	20,000	
10.0		Report to the Court on the District's effort to implement the February, 1982 Order.		
10.1		Prepare December 15, June 15, and July 15 reports to the Court as specified in the amended Section XI of the Court Order.	56,385	
10.2		Conduct process and product evaluation of the Magnets, Academies, and Vanguards. Develop and maintain student and faculty data bases for the Magnets, Academies, and Vanguards. The final report will be included as part of the July 15 report to the Court.	39,870	

(Figure 4. Program Statement: Department of Research, Evaluation, and Audit 1986-87 Continued)

		G	O
10.3	Provide a product evaluation of the programmatic remedies.	18,250	
11.0	Provide statistical information for numerous proposals and reports.		
11.1	Federal and State Proposals	12,500	5,000
11.2	Federal and State Reports	10,150	5,000
11.3	Response to Federal and State Audits	5,000	
12.0	Under the direction of the Administrative Audit Committee, conduct systematic, procedural, and productivity/efficiency reviews to evaluate the efficiency of operations, adequacy of controls, and compliance with Board Policies.		
12.1	Assist external auditors in their annual review of 1985-86 operations as outlined in their engagement letter.	17,750	
12.2	Assist external auditors in their interim review of 1986-87 operations.	22,500	
12.3	Conduct annual audit of activity funds.	17,750	
12.4	Evaluate school controls over financial transactions.	12,500	
12.5	Evaluate the annual physical inventory of service center stock.	5,000	
12.6	Evaluate physical security and controls of input and output in data processing.	30,000	
12.7	Other audits as determined by the Administrative Audit Committee.	59,500	
12.8	Pay external auditor's annual fee.	110,000	

that most evaluation offices spent more time upon summative evaluation and testing than upon any other evaluative activities. Most development work that was undertaken generally involved the development of tests and evaluation instruments rather than that of instructional programs or products. The Webster-Stufflebeam survey, including responses from forty-two of the sixty largest school districts in the United States, tended to support the Lyon data. Product evaluation and testing activities consumed more resources than any other activities. When management and data processing (two activities that support product evaluation and testing) were added, large-sized evaluation units expended 49.6% of their resources in these endeavors; medium-sized units, 58.6%; and small-sized units, 57.8%.

In the ensuing years since 1980, due in part to decreasing allocated resources for evaluation, the previous sentence has become even more true. The golden age of evaluation, from a funding standpoint, is over. The level of formative and summative program evaluation in the nation's public schools has declined precipitously. The testing function remains as a major consumer of resources and a major barometer by which school districts are judged (appropriately or inappropriately).

In fact, today, the testing function consumes a higher percentage of evaluation resources than at any time since the 1960's. In Dallas, just the implementation of the testing program consumes twenty-seven percent of the budget. National data, discussed later in this paper, suggest that this trend is occurring in most districts.

What has basically happened is that the data for program improvement emphasis of the late sixties and seventies has been largely replaced by the accountability movement of the eighties. Whereas, in the late sixties and seventies most school districts were involved in some type of curriculum development and had staffs to do that work; today, the level of curriculum development in school districts has dropped dramatically. Principals, as the managers of school units, are being held accountable for student achievement and consequently are being given more freedom in terms of programs and strategies that they implement in their buildings. When schools fail to produce desired results, principals are often moved instead of programs being abandoned. The evaluation or research emphasis is toward documenting the variables involved in "successful or effective schools"

rather than toward documenting and improving successful programs. Thus, there is a much greater emphasis on criterion-referenced test development to better document effective schools and teachers than on program evaluation designed to improve programs.

This is not to say that there are not a number of competent, well-staffed evaluation units associated with the public schools. There are at least fourteen research and evaluation units across the country that have budgets in excess of one million dollars. The benchmark units of the early and mid-seventies (Atlanta, Cincinnati, Dallas, and Philadelphia) have been joined by large, competent evaluation units in Dade County and Duval County, Florida; Austin and Houston, Texas; Albuquerque, New Mexico; Los Angeles and San Diego, California; Montgomery County, Maryland; New York City, New York; and Washington, D.C. Smaller districts that invest more than ten dollars per student on evaluation include Norfolk, Virginia; Long Beach, California; Mount Diablo, California; Portland, Oregon; Rochester, New York; and St. Paul, Minnesota. In addition, the author is aware of functioning evaluation units in New Orleans, Louisiana; Hillsborough County, Florida; El Paso, Texas; Virginia Beach, Virginia; Denver, Colorado; Milwaukee, Wisconsin; Wichita, Kansas; Corpus Christi, Texas; San Antonio, Texas; Charleston, South Carolina; Indianapolis, Indiana; Oklahoma City, Oklahoma; Boston, Massachusetts; Chicago, Illinois; Cleveland and Columbus, Ohio; Detroit, Michigan; Memphis, Tennessee; Minneapolis, Minnesota; Nashville, Tennessee; Omaha, Nebraska; Pittsburg, Pennsylvania; Seattle, Washington; St. Louis, Missouri; Toledo, Ohio; and Fort Worth, Texas. In fact, ninety-three of the ninety-six largest largest school districts in the United States have organizational units with either research or evaluation in their title.

What do these evaluation units do? In 1978, the author surveyed the sixty largest school systems in the United States regarding their functions. At that time, units were stratified, for analysis purposes, into large, medium, and small. Large units were defined as those that spent more than $1,000,000 on research, evaluation, and testing activities; medium units as those that spent between $300,000 and $999,999 on those activities, and small units as those that spent less than $300,000. Generalizations were then attempted relative to the relationship between size and functions of evaluation departments (Webster and Stufflebeam, 1978).

Utilizing the same questionnaire, the ninety largest school districts in North America were surveyed in 1983. For purposes of analysis, the same stratification was used. Figure 5 defines the relevant functions.

See Figure 5

Figure 6 displays the data relative to the functions of public school research and evaluation units in 1978 and again in 1983. For each activity, the mean, median, mode, and range of statistics within each stratification are reported.

See Figure 6

The most consistent trend in the data reported in Figure 6 relates to the tremendous increase in resources that all units, regardless of size, devote to testing. The large units increased in proportion of resources utilized for testing from 11.8% to 23.5%, the medium from 23.4% to 33.3%, and the small from 19.0% to 30.3%. Instrument development registered an increase in proportion of resources from 3.7% to 10.8% among the large units. These increases, which are consistent with a greater focus on accountability, were accomplished at the expense of most of the other evaluation functions listed. Major casualties, in terms of resources, appeared to be process and product evaluation. Within the context of most evaluation models, process and product evaluation are the major tools of program improvement. Thus, most evaluation units appear to be much more focused on accountability and less on program improvement than they had five years previously.

One other result which is puzzling is the apparent lack of resources that are focused on context evaluation. This probably explains the general lack of longitudinal data bases that appears to be a national phenomenon.

The Future

Many school district research and evaluation units have survived and apparently prospered through some very lean years. Federal funds have continued to dwindle. Despite this predictable cut in Federal resources,

Figure 5. Functions of Research and Evaluation Units

1. Management - the obtaining, disbursing, and quality control of material and personnel resources necessary to meet the information needs of the district.

2. Testing - the operation of the basic system-wide data acquisition system in such a manner as to ensure reliable, valid, and objective instrumentation and reporting. This does not include diagnostic testing for special education.

3. Instrument development - the development and validation of necessary instrumentation.

4. Context evaluation - the provision of baseline information that delineates the environment of interest, describes desired and actual outcomes, identifies unmet needs and unused opportunities, and diagnoses problems that prevent needs from being met (often called institutional research).

5. Input evaluation - the provision of information for determining optimum methods of resource utilization for accomplishing set goals.

6. Process evaluation - the provision of information for determining defects in procedural design or implementation during program implementation stages and for aiding in the interpretation of program evaluation data.

7. Product evaluation - the provision of information for determining the relative success of educational programs. Product evaluation may be interim or summative and generally addresses the extent to which programs meet their objectives, the relative effects of alternative programs, and/or program cost-effectiveness.

8. Applied research - the provision of information pertaining to interactions between student characteristics, teacher characteristics, and/or instructional strategies.

9. Basic research - the provision of information pertaining to fundamental relationships affecting student learning.

10. Data processing - the operation of the basic data retrieval and analysis system.

11. Planning services - the provision of technical assistance to district and project management in planning and managing projects and programs.

12. Ad hoc information - the provision of requested information on an ad hoc basis to decision-makers.

13. Research consultation - the provision of design and analysis help to district personnel in implementing their own research and evaluation projects.

14. Proposal development - the development of proposals, or the evaluation sections of proposals, for outside funds. In all cases outside funds should be a means to an end, not an end in themselves.

Figure 6. Functions of Evaluation Units, 1978 and 1983*

Size of Unit	Activities	management		testing		instrument development		context evaluation		input evaluation		process evaluation		product evaluation	
78 83		78	83	78	83	78	83	78	83	78	83	78	83	78	83
Large	$\bar{X}$	5.8	8.2	11.8	23.5	3.7	10.8	6.8	4.5	3.1	2.4	14.0	3.7	19.0	14.9
N = 9, 12	Me	5	5	12	27	3	5	5	1	3	2	12	3.5	15	10
	Mo	5	5	15	10	1	0	5	0	5	0	10,20,25	0.5	15,20	0.20
	R	3-10	1-20	2-22	1-55	1-12	0-50	2-18	0-15	0-6	0-10	5-25	0-25	6-50	0-60
Medium	$\bar{X}$	8	7.4	23.4	33.3	4.1	5.3	6	2.4	0.2	1.9	10.9	6.3	19.6	10.7
N = 13,18	Me	6	5.5	20	30	4	3.5	6	0.5	0.5	0.5	10.5	5	15.7	10
	Mo	5	3,10	15	20,30	3	0	3	0	0	0	10,20	5	15	10
	R	4-19	2-30	4-70	5-65	2-15	0-20	2-15	0-10	0-3	0-5	1-20	0-25	10-40	2-20
Small	$\bar{X}$	10.7	10	19.0	30.3	4.0	5.7	5.3	5.3	0.6	2.4	5.5	4.2	17.9	10.2
N = 13,13	Me	8	5	15	33	3.5	4.5	5	3.5	1	1	6	5.5	15	10
	Mo	5	5	15	20	5	0	0,10	0	0	0	0,5	5	5	10
	R	5-18	2-60	5-45	0-75	0-10	0-20	0-13	0-25	0-5	0-10	0-10	0-10	5-50	0-20

*Figures represent proportions of research and evaluation budgets.

$\bar{X}$ = Mean
Me = Median
Mo = Mode
R = Range

Figure 6. Functions of Evaluation Units, 1978 and 1983 (Continued)

Size of Unit	Activities	applied research		basic research		data processing		planning services		ad hoc information		research consultation		proposal development	
78 83		78	83	78	83	78	83	78	83	78	83	78	83	78	83
Large	$\bar{X}$	4.0	3.3	2.9	1.0	13.0	4.4	4.6	5.2	3.8	4.0	3.6	2.9	3.9	1.9
N = 9, 12	Me	2	2	1	0.5	10	5	5	4	2	2	3	2	3	1
	Mo	0,2,4	0	0	0	10	5	5	5	2	0	1	0	1,5	0
	R	0-15	0-10	0-18	0-5	5-25	0-11	0-15	0-15	0-15	0-15	0-12	0-15	0-11	0-10
Medium	$\bar{X}$	1.9	4.4	0.5	1.9	7.6	5.2	4.4	7.8	4.4	4.2	4.7	2.6	4.3	1.2
N = 13, 18	Me	3	4.5	0.5	0.5	9.5	10	3.7	4.5	4.3	5.5	2	0.5	3	0
	Mo	1	0	0	0	10	10	4	0	4	5	1,2	0	2	0
	R	0-8	0-10	0-2	0-5	3-16	0-10	2-10	0-20	2-10	0-10	0-10	0.20	1-10	0-15
Small	$\bar{X}$	3.5	2.6	0.3	0.8	10.2	4.5	4.9	6.8	10.8	4.5	4.3	2.5	4.0	1.6
N = 13, 13	Me	5	2.5	0.3	0.2	10	3.5	5	7.5	12	4.5	5	1.5	3.4	1
	Mo	5	0,5	0	0	10	0	5	5,10	5	4	5	0	5	0
	R	0-20	0-6	0-5	0-5	5-15	0-20	1-11	0-20	3-20	0-13	0-8	0-15	0-10	0-10

*Figures represent proportions of research and evaluation budgets.

$\bar{X}$ = Mean
Me = Median
Mo = Mode
R = Range

and a general tightening of tax dollars at the local level, there are at least nineteen public school research and evaluation units from relatively large school districts that spend more than ten dollars per student on research, evaluation, and testing activities. This was not the case in 1978, when the author was able to find only four. Despite inflation, this is a very impressive statistic.

Most of these research and evaluation units apparently have been structured around the concept of accountability. They all appear to spend a relatively large proportion of their respective budgets on testing and test-development activities. Many other evaluation activities have diminished somewhat as a result of this emphasis. It is almost as if superintendents have decided that curriculum people and principals have been refusing guidance from empirical data for too long a time and that now they will be accountable for their product. Evaluation units are being asked to provide reliable and valid data on program and school outcomes. Curriculum people and principals are being held accountable for those outcomes. If those outcomes are less than acceptable, apparently someone else will be given the opportunity to produce acceptable outcomes (hopefully in curriculum, not evaluation).

It is the author's opinion that the accountability movement will gain momentum throughout the remainder of this decade. Evaluators will increasingly be asked to judge the merits of programs and program managers. Personnel evaluation will become increasingly important and, therefore, increasingly systematic. (The vast majority of evaluation units are currently not involved in personnel evaluation.) Evaluation will have an extremely important role in the educational systems of the eighties, but that role will be more oriented toward providing systematic feedback than toward providing suggestions for improvement.

One important concomitant of the increasing pressure for accountability will be a decreasing popularity of evaluators. This will be tolerable as long as a system retains its same superintendent. However, the probability of the head of a successful evaluation department surviving a change in the superintendency is somewhat diminished by this increased role in the area of accountability.

Finally, it is relevant to remember a basic tenet that history repeats itself. The accountability movement will eventually rebreed the basic and applied research movement of the late sixties and early seventies. It is important to remember that much of the success that is currently being enjoyed in public education is built upon the cornerstone of empirically derived knowledge garnered during that earlier period. In short, program evaluation that includes careful process evaluation for implementation documentation, as well as related applied and basic research programs, will return again.

References

Bracht, G.H. The Relationship of Treatment Tasks, Personological Variables, and Dependent Variables to Aptitude-Treatment Interactions. Unpublished doctoral dissertation, University of Colorado, 1969.

Cronbach, L.J. "Course Improvement Through Evaluation", Teachers College Record, 1963, 64, 672-83.

Cronbach, L.J. and Snow, R. Individual differences in learning ability as a function of instructional variables. Final Report, U.S. Officel of Education, Contract No. OEC 4-6-061269-1217. Stanford, California: Stanford University, School of Education, March, 1969.

Lyon, C.D., Doscher, D.L., McGranaham, P., and Williams, R. Evaluation In School Districts, UCLA: Center for the Study of Evaluation, 1978.

Salomon, G.S. "Heuristic Models for the Generation of Aptitude Treatment Interaction Hypotheses", Review of Educational Research, 1972, 42, (3), 327-343.

Scriven, M.S. "The Methodology of Evaluation", in R.E. State (ed.) Curriculum Evaluation, AERA Monograph Series on Curriculum Evaluation, 1, Chicago: Rand McNally, 1967.

Stake, R.E. "The Countenance of Evaluation", Teachers College Record, 68, 1967, 523-40.

Stufflebeam, D. L. "A Depth Study of the Evaluation Requirement", Theory Into Practice, 1966, 5, 121-134.

Stufflebeam, D.L., et al., Educational Evaluation and Decision Making. Itaska, Illinois: Peacock Publisher, Inc., 1971.

Taylor, M.M. and Casserly, M. "The Status of Departments of Research and Evaluation in the Great City Schools", Council of Great City Schools, 1985.

Webster, W.J. "The Dallas Approach to Systematic Research and Evaluation in Development", NCME, New Orleans, Louisiana, February, 1973.

Webster, W. J. "The Organization and Functions of Research and Evaluation in a Large Urban School District", AERA, Washington, D.C., March 31 through April 4, 1975.

Webster, W.J. "Some Observations on the Future of Educational Research and Development in the United States", AERA, Toronto, Ontario, March 27-31, 1978.

Webster, W.J. and Mendro, R.L. "The Investigation of Aptitude-Treatment Interactions as an Integral Part of Program Evaluation." Journal of Experimental Education, Vol. 43, Fall, 1974, pp. 86-91.

Webster, W.J. and Schuhmacher, C.C. "A Unified Strategy for System-Wide Research and Evaluation." Educational Technology, Vol. XIII, No. 5, May, 1973, pp. 68-72.

Webster, W.J. and Stufflebeam, D.L. "The State of Theory and Practice in Educational Evaluation in Large Urban School Districts", invited address presented at the annual meeting of the American Educational Research Association, Toronto, Ontario, March 24-29, 1978.